Application Note

SPORIAN® MICROSYSTEMS, INC

The Use of Machine Learning and High Temperature Raman Spectroscopy for Determining Molten Chloride Salt Chemical Composition

Authors

Duc Nguyen
Kevin F. Harsh
Melissa Mezger
Sporian Microsystems Inc.

Introduction

Molten salts are an important class of fluids for a range of high-temperature applications, including: thermal energy storage and transfer fluids for concentrating solar power (CSP) and molten salt nuclear reactors (MSR), molten carbonate fuel cells (MCFC), electrolysis, metals or waste processing, and ion-exchange (IOX) chemical glass strengthening. This is due to the unique properties of molten salts as stable fluids over a wide range of temperatures (up to 900°C), with low vapor pressures, and with high heat capacities.

Eutectic salt blends have the benefit of very low melting points that are ideal for the most efficient use of molten salts. However, maintaining salt composition control of eutectics can be a challenge. Real-time molten salt composition monitoring is of significant value for molten-salt-based process control. There are few technical options for in-situ, high-temperature-immersed, harsh-environment-tolerant, real-time monitoring that can provide chemical composition information, and most options can only provide partial compositional data.

Raman spectroscopy is a powerful analytical technique that quickly gives highly specific information for the analysis of chemical compounds in a non-destructive manner. Raman-active species exhibit spectra with distinct peaks and provide “fingerprint” information on the vibrational transitions within a molecule by uniquely characterizing

sample volumes in bulk. Raman systems are easily portable and signals can be transmitted by optical fibers over long distances for remote analyses. These features make Raman spectroscopy ideal for implementing a robust, automated molecular identification system.

One drawback of utilizing Raman spectral data for real-time monitoring of molten-salt-based process control is that the data can contain very high dimensionality per sample. Raman-scattered light returning from measuring a sample is dispersed by a diffraction grating and then captured by a Charge Coupled Device (CCD) in a spectrometer. Depending on the length of the CCD array, one measurement of molten salt can contain up to thousands of elements, or pixels, with the spectral information from the sample. Also, depending on the number of salt species targeted for identification and composition quantification, as well as the frequency of measurements during monitoring, the dataset can become very large and complex. Monitoring and analyzing the data by a human analyst would be very limited and time-consuming.

Machine learning (ML), a subset of artificial intelligence, is a data-driven approach to analysis, where instead of fitting a physics-based theory to the data output, an algorithm is used to leverage relationships within the data to generate correlating models. ML is therefore suitable for a situation where the system outputs are unknown, hard to understand, or time-consuming to process by a human analyst. ML models can learn from Raman spectral data, identify patterns, and be able to make decisions with minimal human intervention. The models can rapidly process large amounts of high complexity Raman spectral data to be able to classify unknown chemical compounds and/or monitor the composition in real-time.

In this application note, we'll explore an example of how machine learning was applied to data from a high-temperature Raman spectroscopy monitoring instrument to generate a model that can classify different salt species within a molten salt sample.

Experiment/Methods

Data presented here was collected using a Sporian Microsystems SpecIQ™ High-Temperature Raman Fluid Composition Monitoring System. The Raman spectroscopy-based measurement system has the capability to provide in-situ, real-time monitoring of various molten salts in systems such as those used in CSP, nuclear power, fossil fuel energy storage, materials synthesis/processing, and metal making/finishing. Key system specifications are shown in Table 1, and the system is designed to operate in one of two possible modes: (1) as an autonomous monitoring device that measures and applies machine-learning-based algorithms to provide classification of salt constituents and quantified concentrations to higher-level control systems, and (2) as a user operated instrument with specific feature/function control options.

Table 1: Key SpecIQ™ specifications

Spec/Feature	Unit
Excitation Wavelength	532 nm
Wavenumber Range (Shift)	100-5400 cm^{-1}
Resolution	6 cm^{-1}
Communications	Ethernet
Measurement Temp. Used	25-500°C
Max Measurement Temp. Capable	950°C
Integration Times	0.1-1s

Using the Raman Fluid Composition Monitoring System, we'll explore how data generated from the user operated mode was used to build a machine learning model that could be deployed in the autonomous monitoring mode in the system. A typical machine learning process for Raman spectroscopy involves several stages in order to build a trained model that can be used to analyze unknown sample data[i]. The stages include data

collection, pre-processing, processing, training, validating/evaluating, and deploying the model.

Data collection: In the user-operated mode, data from each of the molten salts were collected in the lab where chloride salts and various eutectic mixtures were melted between 240-900°C. All salt reagents were purchased from Sigma Aldrich with purity ≥ 98%. Mixtures were prepared with reported %wt and ground with a mortar and pestle in a glove box under dry nitrogen to prevent water ingress/absorption by the salts. Salt mixtures were melted in a quartz-lined, environmentally-controlled crucible and heated to the desired temperature with an Ambrell precision induction heating system. The setup included the capability to purge a dry gas through the system during operation. The high-temperature Raman probe hardware was inserted into the heat zone of the system to monitor the mixture when molten, at temperatures ranging from 240 to 900°C. The instrument was configured to measure and send data to a database file that could be subsequently used to process/view resulting data

Pre-processing: The Raman spectral data were 'cleaned' and manipulated to optimize the data. In this stage, data noise was minimized using the measured laser output intensity, uninformative background signals were removed, and spectral intensity values were normalized and scaled. Preparing for hold-out validation of the model is also started at this stage, where the collected data was split into a training group and a test group, 70%/30% respectively. The test group was not seen by the model during training, so it would be tested as unknowns.

Processing: The pre-processed data was transformed and reduced in dimensionality. One common technique includes principal component analysis (PCA), which transforms the spectral features into new coordinate systems that best explain the variance. As a result, the ML model did not need to train on all 2048 elements of the Raman Fluid Composition Monitoring System, and instead analyzed just 20 PCA components of each measurement.

Training: After pre-processing, a classification model was trained on the processed training group using a supervised method. Relationships between the data and the known class labels of the training group were computed by the model. In this example, a support vector classification (SVC) model was implemented. The model is part of the set of support vector machines algorithms, which performs classification by mapping the data in the PCA-transformed components space into a new space separated by a plane that best separate the classes.

Evaluation: Once trained, the classification model was applied to the test group, which was unknown to the model, and predicted the molten salt species in the group. Depending on the results, the model could be tuned by using different parameters in order to improve prediction accuracy or to reduce bias, such as over-fitting.

Deployment: When the model had been trained and sufficiently tuned to optimize performance, both training and test group (all data) were used to train the model. The model was then packaged and deployed into the Raman Fluid Composition Monitoring System. The system could now be deployed into a CSP salt system and engage the autonomous monitoring mode in order to monitor the molten salt in real-time.

Results from each step of the ML process are shown below using an example dataset of chloride salts that were tested at Sporian Microsystems. Scikit-learn's Python library was used for the entire machine learning process[ii].

Results

Figure 1 shows the raw Raman spectral data from the example dataset of chloride salts collected at Sporian. Each salt was scanned 100 times. Since

chloride salts are less Raman active with subtle Raman peak shifts observed below 300 cm^{-1} [iii], the wavenumber range for the dataset was able to be reduced to a much smaller range than the full 5400 cm^{-1} available in the system.

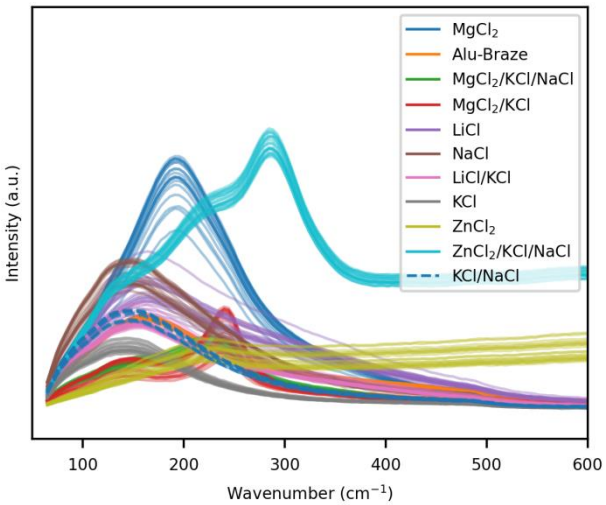

Figure 1: Raw Raman spectra from the example dataset of chloride salts; 100 sample lines per group

The reduced data then went through the pre-processing step of machine learning. After reducing sample-to-sample noise, smoothing, and background removal, the spectral data became much more distinct and more tightly overlapping in Figure 2.

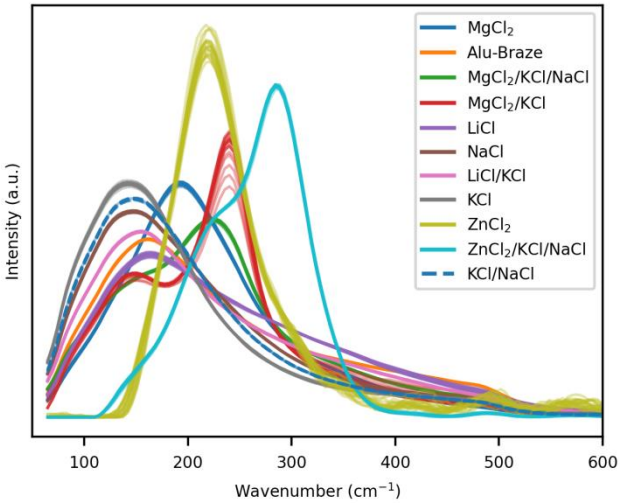

Figure 2: Pre-processed Raman spectral data, including noise reduction, smoothing, and background removal; 100 sample lines per group

As each group contained about 100 sample measurements, these were then separated into 70 training samples and 30 test samples per group. After pre-processing, PCA was applied to further reduce the number of elements. The PCA transform first trained on the training samples only, and then transformed the same training samples. It then transformed the test group using the trained transformation from the training samples. The first three PCA components, which explained >99% of the variance in the data, were plotted in a 3-dimensional figure to show the relationships between the molten salt species. As shown in Figure 3, the tightly overlapping lines from each group transferred well into the PCA transformation. Measurements from each group, represented as single points, formed well-defined clusters that were clearly separated from others.

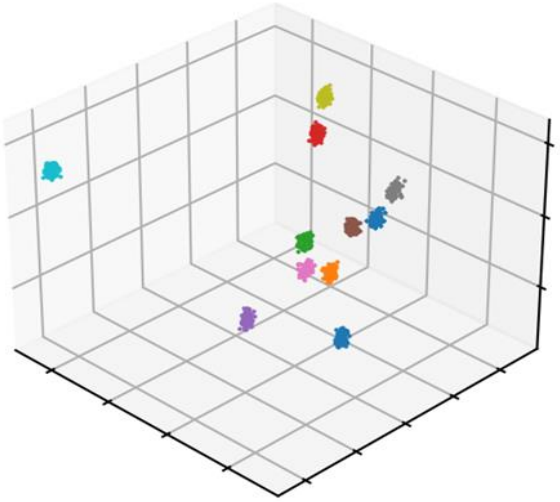

Figure 3: Transformed Raman spectral data shown in the first 3 PCA components

After processing, a classification model used the PCA-transformed training data to train its classification scheme. For this example, a support vector classification (SVC) was used as it was a good fit for data that grouped in clusters. Figure 4 shows the classification results in the form of a Confusion Matrix [iv] when the test data was passed through the SVC model. The Confusion Matrix is a table layout to help visualize the

SPORIAN® MICROSYSTEMS, INC

performance of the classification model. Each row represents the instances from the test group with the actual, true label while each column shows the model's predicted label. Matching rows and columns indicate the model's predictions lined up with the truth. Where the prediction matches with a different row or label, then that incorrect classification is also evident on the Confusion Matrix. These results show that the model was able to accurately predict all of the samples in the test group with 100% accuracy across all molten salt samples.

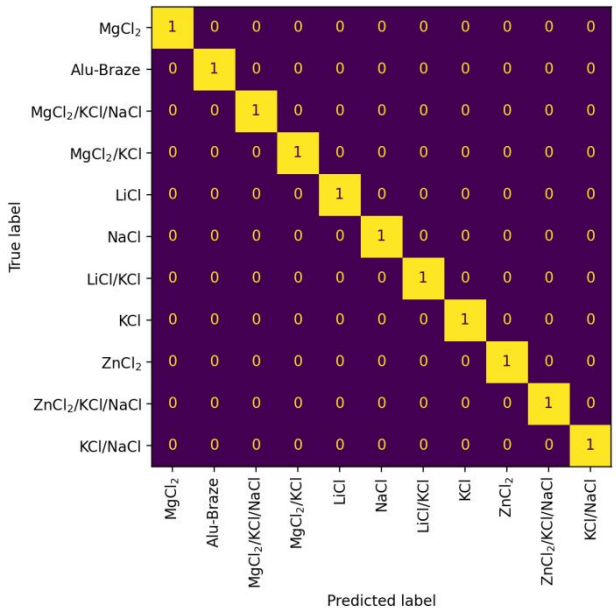

Figure 4: Confusion matrix comparing classification model predictions to their true labels; values show ratio of predicted to true labels, where 1 = 100%

After the model was trained, tested, and verified, all of the available data, including both the training and test groups, can be used to train the classification model. This allows the model to use as much data as possible to build the best model in a real operating scenario. Once fully trained, the model can be deployed in the Sporian Microsystems SpecIQ™ High Temperature Raman Fluid Composition Monitoring System. New data of molten salts in the SpecIQ Raman system can be immediately processed and predictions on the new samples can be obtained in real time. The deployment of the model can be performed remotely over the network, which enables the system to change or update its model as needed without interfering with operations.

Conclusions

Molten salts are a very challenging environment for chemical composition monitoring. Real-time molten salt composition monitoring is of significant value for industries using molten salts. Sporian Microsystems' high-temperature, real-time, Raman spectroscopy-based measurement system is an effective tool to provide in-situ, real-time monitoring of various molten salt systems' compositions, providing constituent and contaminant classification and quantification. Instruments capable of continuous monitoring in very harsh conditions and allowing for the use of machine-learning-based data processing are ideally suited for performing such measurements.

i L. e. a. Pan, "A review of artificial intelligence methods combined with Raman spectroscopy to identify the composition of substances," Journal of Raman Spectroscopy, 2021.

ii F. e. a. Pedregosa, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning research, pp. 2825-2830, 2011.

iii M. H. Booker and C. H. Huang, "Raman spectroscopic studies of structural properties of solid and molten states of the magnesium chloride - alkali metal chloride system," Can J. Chem., vol. 58, pp. 168-179, 1980.

iv Wikipedia contributors. "Confusion matrix." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 23 Sep. 2019. Web. 17 Oct. 2019.